



# Modeling the commute mode share of transit using continuous accessibility to jobs



Andrew Owen <sup>\*</sup>, David M. Levinson

Department of Civil, Environmental, and Geo-Engineering, 500 Pillsbury Drive SE, University of Minnesota, Minneapolis, MN 55408, USA

## ARTICLE INFO

### Article history:

Received 25 October 2013

Received in revised form 29 January 2015

Accepted 9 February 2015

### Keywords:

Public transit

Accessibility

Modeling

Mode share

Mode choice

## ABSTRACT

This paper presents the results of an accessibility-based model of aggregate commute mode share, focusing on the share of transit relative to auto. It demonstrates the use of continuous accessibility – calculated continuously in time, rather than at a single of a few departure times – for the evaluation of transit systems. These accessibility calculations are accomplished using only publicly-available data sources. A binomial logic model is estimated which predicts the likelihood that a commuter will choose transit rather than auto for a commute trip based on aggregate characteristics of the surrounding area. Variables in this model include demographic factors as well as detailed accessibility calculations for both transit and auto. The model achieves a  $\rho^2$  value of 0.597, and analysis of the results suggests that continuous accessibility of transit systems may be a valuable tool for use in modeling and forecasting.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

It is increasingly common for urban transportation planning agencies to establish goals of increasing transit ridership. For example, in 2010 the Metropolitan Council, the metropolitan planning organization for the Minneapolis–Saint Paul, MN area, adopted the goal of doubling transit ridership in its region by 2030 (Metropolitan Council, 2010). In seeking that and similar goals, the council and other agencies will be guided in part by the answer to the question: what makes a traveler choose to make a trip by transit rather than by some other mode?

In almost all cases, that “other mode” is driving: according to estimates from the American Community Survey for the years 2007–2001, 86.5% of commute trips in the Twin Cities metropolitan area were made by car. Driving is a very different experience than using transit: driving is on-demand while transit is schedule-based; a train passenger can read a book while a driver ought not; a bus passenger can talk to other passengers while a solo driver is relatively isolated; a driver may to store her vehicle at the end of a trip while a streetcar passenger pays to board at the beginning of his.

When researchers and planners describe transit and driving as separate *modes* they are focusing on these differences. However, transit systems and road networks share a fundamental purpose: as transportation *systems*, they both are created with the intent that people will use them to reach destinations by paying some cost (in time and/or money). When viewed from this standpoint, it becomes apparent that travel by transit and travel by auto, regardless of their myriad differences, can be compared along two fundamental dimensions: the set of destinations which can be accessed, and the cost of reaching those destinations. Destinations and their cost of access are integrated by measures of accessibility; this study investigated the value of considering the accessibility provided by transit and auto in predicting mode share. In contrast to most existing

<sup>\*</sup> Corresponding author. Tel.: +1 612 624 7550.

E-mail address: [aowen@umn.edu](mailto:aowen@umn.edu) (A. Owen).

accessibility-based models of mode share, the accessibility evaluation method employed here is *continuous* over time – accessibility is evaluated at every minute over the AM peak period, rather than only at a single or a few departure times.

This analysis targets two main research goals:

1. To investigate the feasibility of modeling aggregate commute mode share at origins using detailed accessibility measurement for auto and transit.
2. To test the hypothesis that transit accessibility calculation methods which reflect the ways that transit varies over time provide more accurate models of mode share than methods which use transit accessibility at a single point in time.

The motivation is not to obviate or dismiss the importance of mode share models that rely on detailed demographic data collections, but rather to facilitate transit mode share modeling and forecasting (e.g. sketch planning) in the absence of such data. If an appropriate accessibility metric is able to describe a meaningful proportion of the variation in transit mode share, it can be used to increase the accuracy of current models or decrease the cost of future modeling efforts.

The following section provides a background discussion of accessibility evaluation and mode choice modeling in the context of public transit systems. Subsequent sections describe the data and methodology employed, including a detailed description of the methods used to calculate transit accessibility continuously over time. Finally, the results of a series of mode choice models are presented and discussed.

## 2. Background

### 2.1. Accessibility

Accessibility combines the simpler concept of mobility with an understanding that travel is driven by a desire to reach destinations. It is important to distinguish between *individual accessibility* and *locational accessibility*: the former seeks to characterize the ease with which travelers might reach their destinations, subject to constraints of ability, budget, and other barriers; the latter examines accessibility as a spatial phenomenon by considering the costs and benefits of the potential trips offered by transportation systems between origins and destinations of interest. Horner (2004) explored this distinction in the literature and notes that individual accessibility measures are generally poor at “producing ... generalized assessments of intraurban structure,” while locational accessibility measures are more useful for “understanding relationships between transportation and land use.”

Geurs and Van Wee (2004) provide a taxonomy of accessibility measures and draw a similar distinction between locational and utility-based accessibility measures, and additionally identify *infrastructure-based* measures which focus chiefly on the conditions of a transportation system and only secondarily (if at all) on the origins and destinations served by it. Metrics which indicate congestion or speeds on highway systems, or the headways or service span of a transit system, fall into this category.

Locational accessibility can be a particularly useful tool for transportation planners because it provides a way to evaluate the properties of transportation systems at a level that is aggregate enough to provide guidance for planning at the city and regional level. It can be especially useful for multi-modal transportation planning because it is able to provide a level playing field for evaluating modes relative to one another (Anderson et al., 2012). This is achieved by setting aside the many particular differences between transportation *modes* and considering their relative merits as transportation *systems*.

Many different implementations of locational accessibility measurement are possible. El-Geneidy and Levinson (2006) provide a practical overview of historical and contemporary approaches. Most contemporary implementations can be traced back at least to Hansen (1959), who proposes a measure where potential destinations are weighted by a gravity-based function of their access cost and then summed. The specific weighting function used has a tremendous impact on the resulting accessibility measurements, and the best-performing functions and parameters are generally estimated independently in each study or study area (Ingram, 1971). This makes comparisons between modes, times, and study areas challenging. Levine et al. (2012) discuss these challenges in depth during an inter-metropolitan comparison of accessibility; they find it necessary to estimate weighting parameters separately for each metropolitan area and then implement a second model to estimate a single shared parameter from the populations of each. Geurs and Van Wee (2004) also note the increased complexity introduced by the cost weighting parameter.

Perhaps the simplest approach to evaluating locational accessibility is discussed by Ingram (1971) as well as Morris et al. (1979): cumulative opportunity measures of accessibility employ a binary weighting function. Accessibility is calculated for specific time thresholds and the result is a simple count of destinations that are reachable within each threshold. This analysis employs a cumulative opportunities measure of accessibility following the method described by Owen and Levinson (2012).

### 2.2. Accessibility of transit systems

Lei and Church (2010) provide a review of approaches to evaluating the accessibility provided by transit systems. Developments fall into two categories: changes in the techniques used to calculate travel times by transit, and changes in

the way those travel times are employed to calculate accessibility. The chief technical challenge in evaluations of transit accessibility has been calculating travel times. Prior to the mid-2000s, evaluations of accessibility in transit systems generally operated on simplified representations of transit networks. For example, a bus route might be assigned an average speed, a trip frequency, and hours of service. From these, travel times by transit are estimated rather than measured. Polzin et al. (2002), Beimborn et al. (2003), Wu and Hine (2003), and Shen (2006) follow this general approach. More aggregate evaluations of accessibility, such as those by Kawabata (2003, 2009) and Kawabata and Shen (2007) make use of average travel times reported by transit commuters.

The introduction of the General Transit Feed Specification (GTFS) in 2005 (Google, Inc., 2013) made detailed transit schedules mode widely available, while increases in generally-available computing power made their use more feasible. Krizek et al. (2009), Lei and Church (2010), Benenson et al. (2010), Mavoa et al. (2012), Owen and Levinson (2012) and Dill et al. (2013) demonstrate various calculations of transit accessibility using detailed transit schedules. Despite their technical differences, these studies of transit accessibility are remarkably consistent in the selection and use of departure times to calculate accessibility. In almost every case, the accessibility provided by transit is derived from a single departure time between each origin/destination pair.

Some work has addressed this limitation. Polzin et al. (2002) proposes a “time-of-day-based” evaluation of transit accessibility, and discusses the fact that transit service levels vary throughout the day. However, the ultimate focus is on variation in demand: after calculating accessibility on a simple hypothetical two-route transit network, the results are scaled based on the distribution of passenger trips throughout the day. Mavoa et al. (2012) addresses the issue of accessibility variation by reporting a transit frequency measure alongside the accessibility value for each analysis zone. However, the accessibility values themselves are based on travel times calculated at a single departure time. Similarly, Dill et al. (2013) include a single-departure-time accessibility variable when modeling transit ridership in addition to nine other variables describing local service levels.

Lei and Church (2010) propose a method for evaluating transit accessibility that is sensitive to travel time variations throughout the day. This approach calculates accessibility by using detailed schedule information to find the minimum travel time in an arbitrary trip departure window. Owen and Levinson (2012) follow a similar approach, guided by the earlier work of Krizek et al. (2009). While this makes the selection of a departure time less arbitrary, it still makes the assumption that transit users are willing and able to adjust their departure time, within an arbitrary window, in order to achieve this optimal travel time.

Fan et al. (2012) provide the clearest example of how transit accessibility can be evaluated across multiple departure times. Accessibility values are calculated using travel times based on departures at each hour of the day; these are averaged to produce a single accessibility metric which incorporates travel times at multiple departure times.

### 2.3. Continuous transit accessibility

Anderson et al. (2013) propose a method for implementing a measurement of transit accessibility that captures the way that accessibility fluctuates continuously over time as trips approach and depart. For each trip departure  $n$ , the departure time and a vector  $T_n$  representing the travel times provided by that trip to all reachable destinations  $m$  are retained. A vector  $O$  is also established which provides the number of opportunities at each reachable destination:

$$T_n = \langle t_1, t_2, \dots, t_m \rangle$$

$$O = \langle o_1, o_2, \dots, o_m \rangle$$

Using the selected accessibility function, the accessibility provided at the departure time for each trip is calculated. A time sampling interval is then selected and the calculation moves backwards through time from each trip departure time, applying the same accessibility function to the next trip's travel time vector with the current time offset subtracted from each element. When the departure time  $d$  for the previous trip  $n$  is reached, the process is restarted using the travel time vector for that trip is the accessibility provided by the previous trip is greater than the accessibility provided by waiting for the next trip. Thus it is possible to calculate accessibility  $A$  at each time point  $t$ :

$$A_t = \begin{cases} A_{T_1 + \Delta t_1, 0} & \text{if } t \leq d_1 \\ A_{T_2 + \Delta t_2, 0} & \text{if } d_1 < t \leq d_2 \\ \dots & \\ A_{T_n + \Delta t_n, 0} & \text{if } d_{n-1} < t \leq d_n \end{cases}$$

Once accessibility is calculated at every time point of interest, it can be treated as a continuous variable over time.

### 2.4. Models of commute mode choice, mode share, and transit ridership

McFadden's work of the 1970s and 1980s explored the theories of discrete choice and demonstrated the application of logit models to transportation mode choice (1974a, 1974b). Mode choice is estimated at the individual level using single estimated travel times via auto and transit for each individual's home-work trip (1973). Today, it is possible to calculate in far

greater detail the benefits provided by entire transportation systems, and to evaluate all possible trips from a given origin, rather than only the trips actually taken by a set of survey respondents.

Mode share within an analysis zone summarizes the result of local individuals' mode choices. Research into mode share and mode choice has explored a wide range of factors which potentially contribute to individuals' mode choice process. Taylor et al. (2009) provide a review of prior research into transit mode share which identifies two major avenues of investigation: *descriptive* analyses which focus on "traveller attitudes and perceptions," and *causal* analyses which examine "environmental, system, and behavioral characteristics." Within each, individual factors are identified as *external* if they are generally outside the direct control of transit planners and managers (e.g. population, income), or *internal* when they are endogenous to a specific transit system implementation (e.g. fares, vehicle design).

Using this classification, accessibility-based investigations of transit mode share can be described as causal, because they specifically avoid reliance on demographics or traveller preferences. They included both internal and external factors: travel times are the direct result of routing and scheduling decisions made by the operating agency, while the spatial distributions of opportunities are not.

Dill et al. (2013) provide an excellent example of contemporary transit ridership modeling as well as a comparison of three transit ridership models implemented by transit agencies in the Portland, OR region. The best-performing model achieves an adjusted  $R^2$  of 0.69 using twenty-nine independent variables categorized as "socio-demographic," "transit service," "transportation infrastructure," or "land use" variables.

Lei and Church (2010) discuss the use of accessibility measurement to make comparisons between different transportation modes, pointing out that service level indicators alone cannot facilitate direct comparisons between modes. Church and Marston (2003) specifically argue for the use of relative accessibility as a tool to understand decision-making in transportation. Benenson et al. (2010) propose specific methods for comparing transit and auto accessibility, with the goal of identifying "accessibility gaps" where the difference between the accessibility levels of the two systems is greater than expected or desired.

## 2.5. Summary

In summary, accessibility evaluation of transit systems is an area of active exploration, and researchers are aggressively seeking new ways to exploit improvements in available data and computation resources. Recent work has focused in particular on making accessibility evaluations more sensitive to variation in transit service frequency over time, and on using accessibility metrics as a tool for comparisons between transportation modes.

## 3. Data

Any analysis involving accessibility requires data describing the locations of origins and destinations, the cost of traveling between them, and the opportunities available at each.

### 3.1. Auto

Automobiles travel across the network of public roads and highways. Calculating travel times through this network requires two types of information: data describing the structure of the network, and data describing the cost of travel along individual links in the network. The Metropolitan Council maintains a model of the regional road network. It provides a network topography for freeway, arterial, and collector roadways in the region, designed to model travel using transportation analysis zones (TAZs) as origins and destinations.

By itself, this model only describes the structure of the road network; per-link speed information is needed in order to generate travel times. The Twin Cities' regional freeway network is very well-instrumented, and data recorded by loop detectors throughout the system are archived by the Minnesota Traffic Observatory, operated and hosted by the University of Minnesota. Archived loop detector data for every weekday in 2010 provided the basis for average freeway link speeds.

Freeway speeds are derived from direct traffic observations made by embedded loop detectors which record the observed traffic volume and detector occupancy at 30-s intervals. MnDOT provides an estimated average effective vehicle length for each detector, which allows the calculation of speed from volume and occupancy. These 30-s speed measurements are aggregated by averaging to 5-min time slots. Finally, the 5-min speed measurements taken between 7 and 9 AM on weekdays during the year 2010 are averaged for each detector to produce a representation of average AM peak period speeds. The resulting detector speeds are assigned to links in the model network based on location.

In contrast to freeways, local arterials and collectors are only sparsely instrumented. Average arterial and collector link speeds are estimated from speed measurements made during a regional GPS-based travel survey in 2008 (Zhu, 2010). This data represents a very accurate measurement of traffic speeds at specific locations and specific times. Speeds for unobserved links are estimated from the samples collected on links with a similar functional classification.

This approach to road speed measurement and estimation can be expensive and labor-intensive. In the time since this speed evaluation was performed, other work has demonstrated alternatives which can reduce the cost and effort associated with collecting data for auto travel times by relying on online trip planning services such as those provided by Google, Inc.

Wang and Xu (2011) provide an example of this approach and demonstrate its use in an evaluation of accessibility to health care facilities.

### 3.2. Transit

Transit users interact with a different type of network than automobile users. Instead of navigating physical infrastructure, transit users move through a more abstract network of bus and rail routes provided by the transit operator. Metro Transit, the primary transit operator in the Twin Cities region, provided a publicly-available GTFS dataset. The schedule used in this analysis was published on November 9, 2009 and describes transit services in operation through March of 2010.

Transit travel time calculations include off-vehicle time costs: waiting at stations as well as time spent accessing an initial station, accessing any required transfers, and accessing the destination after disembarking. This requires a detailed representation of pedestrian facilities in order to calculate walk times between origin and destination census block and transit stations. OpenStreetMap provides an open-source dataset with sufficient detail for this purpose. Specifically, the pedestrian network is comprised of OpenStreetMap feature with the “footway,” “pedestrian,” and “residential” tags.

### 3.3. Labor and employment

Data describing the distribution of labor and employment in the region are drawn from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) program. The workplace area characteristics dataset for 2010 provides census block-level estimates of employee home and work locations.

In general, LEHD is a useful source for accessibility evaluation because it is updated yearly and is drawn from actual payroll records collected at the state level – in this case, by the Minnesota Department of Employment and Economic Development. However, it is important to recognize the fact that LEHD data are *synthetic*: while they are based on actual payroll records, the published results are created by an algorithm designed to produce data which are statistically similar to the underlying data, and which converge to the same distribution when aggregated. An analysis by Spear (2011) of LEHD data in transportation analysis found LEHD to be a useful source of both home and work location data, but identified shortcomings related to job locations of federal workers. Tilahun and Levinson (2011), Fan et al. (2012) and Horner and Schleith (2012) demonstrate the use of LEHD data in contemporary transportation research; the latter also compare LEHD to Census Transportation Planning Products data, which provide similar information at a more aggregate resolution. A Census Bureau publication (Graham et al., 2014) highlights complementarities between LEHD and ACS data, and provides an overview of LEHD data collection and processing methodology.

### 3.4. Mode share

The American Community Survey (ACS) collects data describing commute mode choice during its annual national survey of households and individuals. This analysis uses the 2007–2011 5-year estimate ACS dataset. For each of the 2085 block groups in the study area, the ACS estimated the total number of workers, which includes the workers who report commuting by each mode as well as workers who report working from home.

Because of the specific methodology used to calculate transit accessibility in this analysis (described in Section 4.1), transit and walking are considered a single mode. All modes that involve motorized vehicles on public roads are grouped as “Auto.” In order to model a binary mode choice workers who commute via bicycle, who work from home, or who commute via some other means are excluded. Fig. 1 illustrates transit mode share in the study area.

## 4. Methodology

### 4.1. Calculating travel times by transit

Travel times by transit are calculated using OpenTripPlanner, an open-source graph-based transit routing system which operates on a unified graph including links representing road, pedestrian, and transit facilities and services. For each origin, travel times to surrounding destinations are calculated repeatedly, using each minute between 7 and 9 AM as a departure time. This approach builds on the work of Anderson et al. (2013) by adopting the convention of minute-by-minute travel time calculations and implementing it using existing open-source transit routing software. Compared to previous departure-time-sensitive evaluations of transit accessibility, it provides a much higher temporal resolution than hourly departure times (e.g. Fan et al. (2012)) or at a single, optimal departure time for each origin (e.g. Lei and Church (2010), Owen and Levinson (2012)).

The time cost of travel by transit is comprised of several components. *Initial access time* refers to the time cost of traveling from the origin to a transit stop or station. *Initial wait time* refers to the time spent after reaching the transit station but before the trip departs. *On-vehicle time* refers to time spent on board a transit vehicle. When transfers are involved, *transfer access time* refers to time spent accessing a secondary transit station and waiting there for the connecting trip. Finally, *destination access time* refers to time spent traveling from the final transit station to the destination. All access portions of the



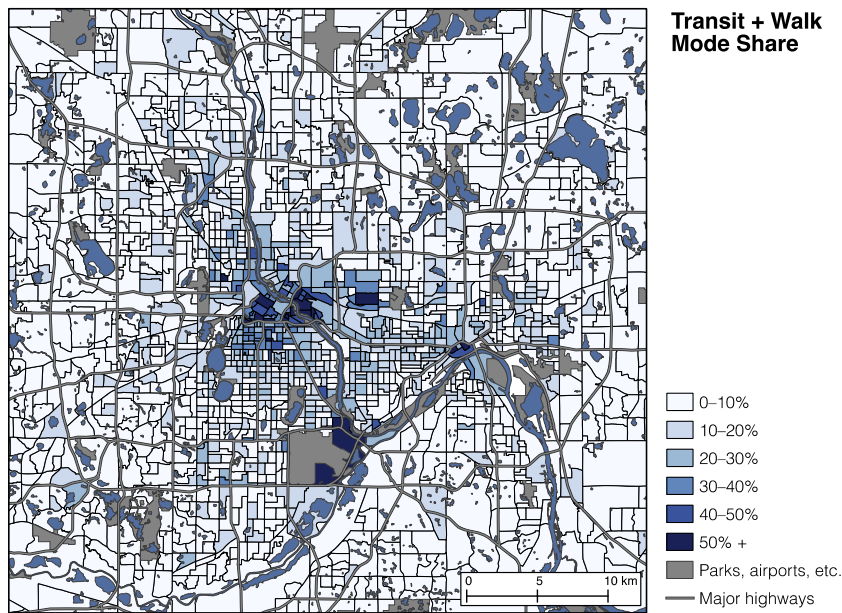


Fig. 1. Commute mode share of transit + walking (ACS 2007–2011 estimates).

trip take place by walking at a speed of 1.38 m/s along designated pedestrian facilities such as sidewalks, and trails. On-vehicle travel time is derived directly from published transit timetables, under an assumption of perfect schedule adherence.

An unlimited number of transfers are allowed. This is somewhat unusual among evaluations of transit accessibility. In many cases travel times are limited to trips involving no more than one or two transfers; this is justified by the observation that in most cities a very large majority (often over 90%) of observed transit trips involve no more than two transfers. However, the shortest-path algorithms typically employed in these evaluations are single-constraint algorithms: they are guaranteed to find the shortest path only when given a single constraint. When the path search tree is pruned based on an additional constraint such as number of transfers, these algorithms provide no insurance against a shorter trip, requiring additional transfers, remaining undiscovered in the pruned space (Korkmaz and Krunz, 2001; Kuipers et al., 2002).

Given the realities of transit networks, it is likely that cases where (for example) a three-transfer itinerary provides a faster trip than a two-transfer itinerary are rare. However, given the goal of evaluating the full accessibility provided by a transit system rather than simply the accessibility that is likely to be utilized, this analysis prefers the algorithmically correct approach of using travel times as the single routing constraint and leaving the number of transfers unconstrained.

Transit and walking are considered to be a single combined mode. The practical implication of this is that the shortest path by “transit” is not required to include a transit vehicle. The shortest walking path from an origin to a transit station in some cases passes through potential destinations where job opportunities exist. In other cases, the shortest walking path from an origin to a destination might pass through a transit access point which provides no trips which would reduce the origin–destination travel time. In these situations, enforcing a minimum number of transit boardings would artificially inflate the shortest-path travel times.

It is difficult to evaluate how this “transit + walking” approach relates to previous work. While a significant majority of the studies discussed in Section 2 use walking as the transit access mode and include walking components in transit travel time calculation, none specify whether walking-only access to destinations is allowed. Because the vast majority of actual transit trips do include walking components, this approach may theoretically improve model performance by more completely including the opportunities that can be reached within a given time budget.

#### 4.2. Calculating travel times by auto

Travel times by auto are calculated using a shortest-path search on a graph defined by the Metropolitan Council’s 2009 regional road network model and the speed data sources discussed in Section 3.1 above. It should be noted that auto travel times do not include an estimation of access times at the origin (e.g. walking from home to a car) or at the destination (e.g. walking from a parking space to work). While the former is likely negligible given the typical residential urban forms in the study area, the latter has the potential to be significant. Unfortunately, the spatial resolution of the road and highway network model available for this analysis is not detailed enough to allow accurate estimation of destination access times.

### 4.3. Calculating accessibility to jobs

Using the travel times for auto and transit described in Sections 4.1 and 4.2, cumulative opportunities accessibility to jobs is calculated for each origin using values of 10, 20, 30, 40, 50, and 60 min for the time threshold ( $t$ ). For auto accessibility, this takes place using TAZs as origins and destinations; for transit, census blocks are used as origins and destinations. Later, the geographic resolution of all accessibility data is standardized as described in Section 4.5.

### 4.4. Calculating time-continuous transit accessibility

Transit accessibility to jobs is evaluated using every minute in the 7–9 AM peak period as a potential departure time. Fig. 2 illustrates how accessibility varies minute by minute at a single census block. Accessibility increases as departure times at nearby stops approach, and then drops after trips depart. Deep troughs in the accessibility profile are associated with times with few or no upcoming trip departures at nearby stops, while sustained periods of high accessibility are associated with periods providing frequent departures. Because of these fluctuations, the average accessibility over the 7–9 AM peak period is significantly lower than the maximum accessibility value over the same period.

Fig. 3 provides a map of average accessibility at the block group level over the 7–9 AM peak period for each block group, while Fig. 4 illustrates maximum accessibility over the same period. Average accessibility values are across the board lower than the maximum accessibility values, but the magnitude of the difference between the maximum and the average accessibility values varies throughout the region.

In addition to average accessibility, calculation of time-continuous transit accessibility allows computation of various measures of variance. For each block group, the standard deviation, variance, and coefficient of variation are calculated over the 7–9 AM peak period. As shown in Fig. 5, transit accessibility varies more widely in some parts of the metro area than others. In central Minneapolis, downtown Saint Paul and a few places in between, very low accessibility coefficients of variation suggest that high-frequency transit service is preventing gaps between trip departures where accessibility would otherwise drop.

In contrast, in the inner-ring suburbs, especially to the west and northwest of Minneapolis as well as to the south of Saint Paul, the accessibility coefficients of variation tend to be very high. This is not necessarily a negative phenomenon; it merely indicates a different approach to transit service. Transit routes in these areas tend to be fairly infrequent, running every half hour or hour. Between these trips accessibility is very low. Additionally, these areas often have express commuter service during the AM and PM peak periods, which provides very high accessibility per trip but remains low-frequency. Transit accessibility in these areas varies sharply over time, and the local coefficients of variation reflect that.

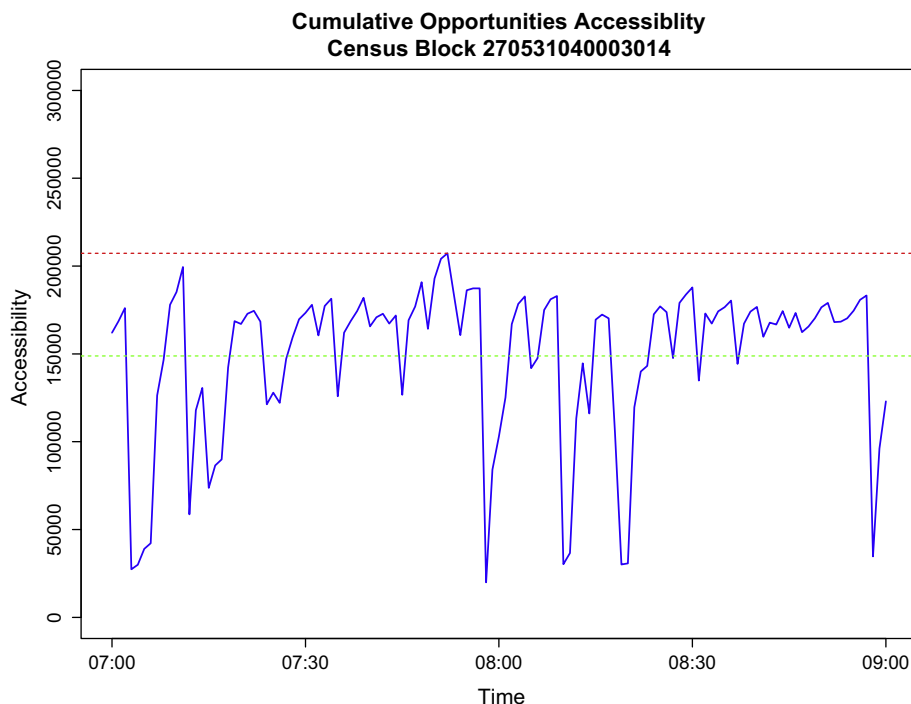


Fig. 2. Time-continuous transit accessibility for a single Census block over the 7–9 AM period.

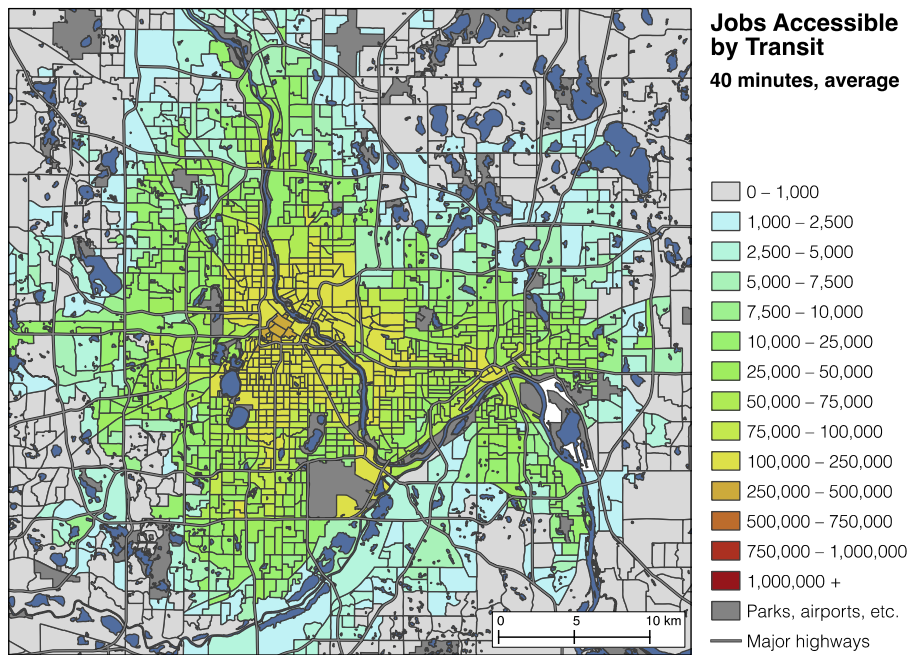


Fig. 3. Jobs accessible within 40 min by transit, 7–9 AM (average).

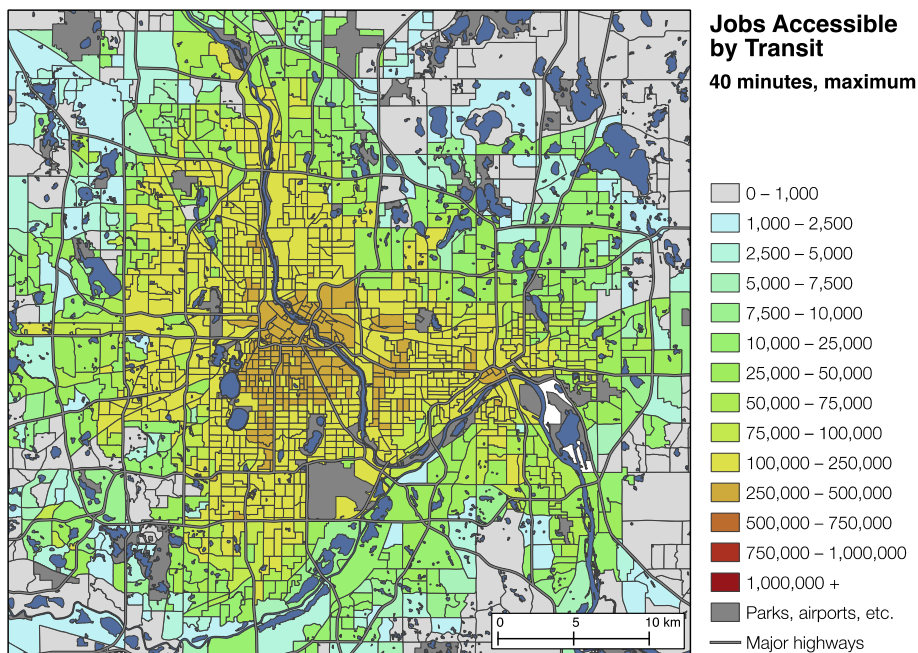


Fig. 4. Jobs accessible within 40 min by transit, 7–9 AM (maximum).

#### 4.5. Geographic standardization

The three principal inputs to this analysis are available at three different levels of geographic resolution. These are standardized in order to use them in combination.

Auto accessibility values are assigned to block groups based on centroid inclusion. Each block group is assigned the accessibility of the TAZ which includes its centroid. Because there are roughly twice as many block groups in the Twin Cities metropolitan area as there are TAZs, neighboring block groups sometimes receive the same auto accessibility value.



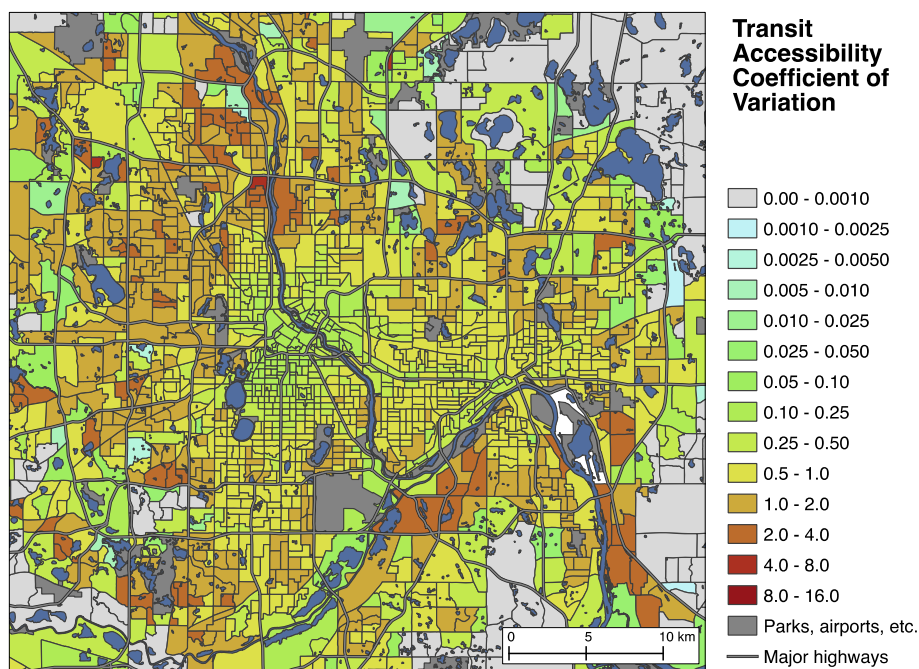


Fig. 5. Transit accessibility coefficient of variation over 7–9 AM period.

Transit accessibility is aggregated to the block group using worker-weighted averaging. Within each block group, the accessibility values for each contained block are assigned weights proportional to the local number of resident workers, and then averaged using these weights. Thus the final accessibility value for each block group represents the accessibility experienced by the average working living within it.

Because this analysis used data originally collected at a variety of geographic scales and zone definitions, it may be particularly sensitive to concerns related to the modifiable areal unit problem (MAUP). The MAUP is discussed in the context of commuting behavior research by Horner and Murray (2002) as well as Niedzielski et al. (2013). In particular, it is important to note that for all three of the zone types used (census blocks, block groups, and TAZs), scale varies in a spatial pattern closely matching that of the accessibility data: central areas have small zones and high accessibility, while outlying areas have large zones and lower accessibility. This is particularly true for auto accessibility patterns. Additionally, the auto accessibility data uses the most aggregate zone type (TAZs), while the preferred MAUP countermeasure is to use data sources as disaggregate as possible. Overall, these concerns suggest that the auto accessibility may be the weakest data component of this analysis.

#### 4.6. Model formulation

Transit accessibility is considered to be the aggregate outcome of individual commuters' mode choices. The choice modeled here is the choice between (1) using transit or walking, and (2) using an automobile or an auto-like mode. These will subsequently be referred to simply as "transit" and "auto." This choice is modeled using the binomial logic form, where the likelihood of an individual making each choice depends on the utility indicated by a specified set of variables. Transit is always regarded a "choice 1," so a positive coefficient will always indicate a positive relationship with the probability of using transit. The following sections describe the various model formulations tested in this analysis.

Models 1 through 4 use independent variables derived only from evaluations of the transit and auto transportation systems. They explore the relationships between transit mode share and various combinations of average transit accessibility, maximum transit accessibility, and variance of transit accessibility. Various combinations of these variables are tested to discover which provides the best-performing representation of the utility provided by the transit system. Auto accessibility is included in all models as a representation of the utility provided by driving. Models are estimated using groups of parameters for each 10-min accessibility threshold increment; the same threshold is used for all parameters in each model. The best-performing model is extended in model 5 to include variables derived from socioeconomic and demographic data. Table 1 describes each variable.

Additionally, each model includes the estimation of an alternative-specific constant (ASC) to capture the variation in transit mode share that is not explained by the above parameters.

**Table 1**  
Variable descriptions and statistics.

Variable	Description	Mean <sup>a</sup>	S.D. <sup>a</sup>
$A_{Tm}$	Maximum transit accessibility (100,000 jobs)	0.874 <sup>b</sup>	0.995 <sup>b</sup>
$A_{Ta}$	Average transit accessibility (100,000 jobs)	0.358 <sup>b</sup>	0.557 <sup>b</sup>
$A_{Tv}$	Transit accessibility variance (100,000 jobs)	8.449 <sup>b</sup>	12.06 <sup>b</sup>
$A_A$	Auto accessibility (100,000 jobs)	13.48 <sup>b</sup>	2.95 <sup>b</sup>
$I$	Mean household income (\$1000)	69.89	30.64
$S$	Mean household size	2.51	0.53
$V$	Mean vehicles per household	1.77	0.45
$W$	Percent white, non-Hispanic population	81.3	22.89
$B$	Percent of population 25 + with B.A./B.S. or higher	25.7	11.92

<sup>a</sup> Over 2082 block groups.<sup>b</sup> 40-min threshold.

#### 4.7. Model evaluation

In order to evaluate the several models that will be presented, it will be necessary to measure two things. First, it must be possible to evaluate how well each model fits the data. Second, it must be possible to evaluate how each model performs relative to the others.

Model fit is evaluated using the pseudo- $R^2$  for logistic regression labeled as  $\rho^2$  by McFadden (1974a, 1974b), and calculated from the likelihood measures of the specific model and the null model. In an evaluation of pseudo- $R^2$  measures in the context of logistic regression, Menard (2000) identifies several desirable qualities of  $\rho^2$ . Most importantly, it was found to have an “intuitively reasonable interpretation as a proportional reduction in error measure.”

Relative model performance is evaluated using the Akaike information criterion (AIC). AIC estimates the amount of information lost by using the model instead of the data, and penalizes models for including more parameters (Burnham, 2002). The latter is an important property given that this analysis depends on comparisons between models with different numbers of parameters.

### 5. Results and discussions

Table 2 summarizes the relative performance of the various models at each accessibility time threshold, as measured by the AIC of each model. Lower AIC values indicate more information preserved in the model, and thus greater model performance. Higher  $\rho^2$  values indicate a better fit between the model and the data.

A comparison of models 1 and 2 indicates that at time thresholds below 30 min, maximum transit accessibility provides a better-performing model than does average transit accessibility, but the opposite is true at thresholds of 30 min and higher. However, at time thresholds below 30 min, model performance is relatively low regardless of what parameters are included. The most significant improvements in model performance appear when moving from a 20-min to a 30-min accessibility threshold.

When combined in model 3, maximum and average transit accessibility together provide a consistently better fit for the data than either does individually, with model fit peaking at the 40-min threshold. Across all time thresholds, the addition of variance of accessibility in model 4 provides a small but statistically significant improvement in model fit. Model 4 at a 40-min threshold provides both the best fit of the transportation-based models, as indicated by a  $\rho^2$  of 0.545, and the best preservation of information relative to the data, as indicated by an AIC of 68858. The introduction of socioeconomic and demographic variables in model 5 improves the model fit at all thresholds. The difference in performance between model 5 and any of models 1 through 4 is much greater than between any of those models individually.

All models decrease in performance when moving from a 10-min threshold to a 20-min threshold; this is an unexpected result. One possible explanation lies with the fact that the transit accessibility calculation method allows pure walking trips as well. If, as seems possible, only a small share of transit commute trips are shorter than 30 min while only a small share of

**Table 2**  
Model performance using various time thresholds.

Threshold	Model 1		Model 2		Model 3		Model 4		Model 5	
	AIC	$\rho^2$	AIC	$\rho^2$	AIC	$\rho^2$	AIC	$\rho^2$	AIC	$\rho^2$
10	74,786	0.506	75,027	0.504	74,786	0.506	73,952	0.512	63,091	0.583
20	77,309	0.489	79,606	0.474	77,297	0.490	76,198	0.497	65,436	0.568
30	74,043	0.511	72,547	0.521	71,012	0.531	70,657	0.533	62,422	0.588
40	75,272	0.502	69,783	0.539	69,460	0.541	68,858	0.545	61,219	0.596
50	76,990	0.491	70,257	0.536	70,179	0.537	69,682	0.540	61,097	0.597
60	78,786	0.480	70,676	0.533	70,672	0.533	70,080	0.537	61,151	0.596

**Table 3**

Commute mode share model results.

Variable	Model 1		Model 2		Model 3		Model 4		Model 5	
	Value	z	Value	z	Value	z	Value	z	Value	z
$A_{Tm}$	0.963	181.9***			0.152	18.1***	0.423	30.7***	0.008	0.7
$A_{Ta}$			1.100	235.1***	0.902	74.7***	0.669	44.0***	0.652	49.6***
$A_{Tv}$							−0.013	−24.4***	−0.002	−5.4***
$A_A$	0.055	46.5***	0.089	53.3***	0.075	71.9***	0.068	37.5***	0.052	16.1***
$I$									−0.004	−19.6***
$S$									0.250	29.7***
$V$									−0.835	−50.2***
$W$									−0.117	−5.1***
$B$									0.596	15.5***
ASC	−3.54	−298.7***	−4.14	−182.8***	−4.01	−173.4***	−3.95	−171.6***	−2.63	−45.6***
$\rho^2$	0.511		0.539		0.541		0.545		0.597	
AIC	74,043		69,783		69,460		68,858		61,097	
Threshold	30		40		40		40		50	
$n$	2082									

\*\*\* Significant at 0.001 level or better.

walking commute trips are longer than 20 min, it may be that the 20-min transit/walk accessibility does not align well with actual traveler behavior.

Table 3 lists the coefficients, and their significant, as estimated in the various models. In general, the accessibility coefficients follow expected and intuitive patterns. Increases in both maximum and average accessibility are associated with increases in transit mode share, while increased variance of transit accessibility is associated with lower transit mode share.

However, the coefficients for the auto accessibility parameter ( $A_A$ ) are unexpectedly positive – indicating that higher auto accessibility is associated with higher transit mode share. As illustrated in Table 1, auto accessibility is, throughout the study area, much higher than transit accessibility – often by multiple orders of magnitude. Also, the higher speeds associated with the road network mean that auto accessibility varies far more gradually over space than does transit accessibility, which peaks at locations near transit stops and stations but then rapidly drops off. As discussed in Section 4.5, the auto accessibility data used here is likely the component most susceptible to confusion caused by the MAUP. Combined, these suggest that meaningful comparisons of accessibility between transit and auto may require more careful investigation of the scale and scope of the variation. It may be useful to limit the analysis to areas which have some minimum level of transit accessibility or which are within some distance of transit service, thereby excluding large outlying zones where transit accessibility is effectively zero and there is little variation in auto accessibility. Additionally, it would be worthwhile to explore methods for calculating auto accessibility that can provide more detailed and disaggregate data.

In model 5, the signs and magnitudes of coefficients of the sociodemographic variables generally follow the patterns suggested by previous literature. Transit mode share is negatively associated with household income ( $I$ ), a relationship that is found in nearly all investigations of transit ridership in North America. Some of the demographic variable results are also expected: transit mode share has a negative association with increasing share of white, non-Hispanic population ( $W$ ), and a negative association with increasing ownership of private vehicles ( $V$ ). However, the coefficients for household size ( $S$ ) and education ( $B$ , % with bachelor's degrees or higher) are unexpected. Dill et al. (2013) found negative associations between these variables and transit ridership in Portland, OR. Since these studies represent only two metropolitan areas, further research is necessary to establish the implications of these potentially contradictory results.

Also, it is important to note that in this model the effect of maximum transit accessibility is not statistically significant, though both average transit accessibility and variance of transit accessibility maintain their significance. This unexpected finding regarding maximum transit accessibility may be related to factors such as residential location selection or work arrival time flexibility, both of which have logical associations with the demographic variables included in the model.

Finally, this aggregate approach to mode share modeling requires careful interpretation. It is most important to recognize that the models' dependent variable is mode share (not mode choice), and that the models are based on aggregate zonal characteristics, rather than individual characteristics. As a result, the finding that (for example) a higher median income is associated with lower transit mode share does not necessarily imply that an individual with a higher income is less likely to commute using transit. More detailed, multi-level modeling approaches can make more specific inferences about individuals' behavior.

## 6. Conclusion

Thanks to improvements in data sources and processing capabilities, detailed accessibility calculations are practical today which would have been previously unachievable. This analysis has built on earlier work which established methods for calculating continuous transit accessibility in two primary ways.

First, it has demonstrated the feasibility of accessibility-based mode share modeling. Though questions remain as to the most appropriate model specification, accessibility-based modeling can provide useful results even in the absence of socioeconomic and demographic data. Which such data are added to an accessibility-based model, accessibility parameters retain a statistical and practical significance. Second, it has demonstrated the value of accessibility evaluation methods for transit systems which reflect the variation of accessibility over small time scales. Though most accessibility-based investigations of transit use evaluate accessibility only at a single departure time, this research shows that model fit may be improved by incorporating time-averaged accessibility and/or measures of accessibility variation over times.

This work also highlights existing challenges in using accessibility as a tool for cross-modal comparisons (in this case, between transit and auto). It is rare to find data describing these two transportation systems that are of consistent scale and scope, both spatial and temporal. As a result, it is necessary to make compromises in data processing that may increase the error associated with the modifiable areal unit problem. Continued improvements in the quality and availability of high-resolution data sources may help mitigate this problem. Finally, there is a need to further test the robustness of this approach through application to multiple geographic areas.

Overall, these results suggest that while the specific model formulations presented here can be improved upon, continuous evaluations of the accessibility provided by transit systems are a promising metric for use in ridership and mode share modeling. The techniques demonstrated allow detailed calculations of transit accessibility to be implemented at low cost and with data that are generally publically available. It may be possible to apply these techniques to more robust existing models of transit ridership and mode share to improve their performance.

## References

- Anderson, P., Levinson, D.M., Parthasarathi, P., 2012. Using Twin Cities destinations and their accessibility as a multimodal planning tool. Minnesota Department of Transportation.
- Anderson, P., Owen, A., Levinson, D., 2013. The time between: continuously-defined accessibility functions for scheduled transportation systems. In: 92nd Annual Meeting of the Transportation Research Board.
- Beimborn, E.A., Greenwald, M.J., Jin, X., 2003. Accessibility, connectivity, and captivity: impacts on transit choice. *Transport. Res. Rec.: J. Transport. Res. Board* 1835 (1), 1–9.
- Benenson, I., Martens, K., Rofé, Y., 2010. Measuring the gap between car and transit accessibility. *Transport. Res. Rec.: J. Transport. Res. Board* 2144 (1), 28–35.
- Burnham, K.P., Anderson, D.R., 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, second ed. Springer-Verlag.
- Church, R.L., Marston, J.R., 2003. Measuring accessibility for people with a disability. *Geogr. Anal.* 35 (1), 83–96.
- Dill, J., Schlossberg, M., Ma, L., Meyer, C., 2013. Predicting transit ridership at the stop level: the role of service and urban form. In: 92nd Annual Meeting of the Transportation Research Board.
- El-Geneidy, A.M., Levinson, D.M., 2006. Access to destinations: development of accessibility measures. Minnesota Department of Transportation.
- Fan, Y., Guthrie, A., Levinson, D., 2012. Impact of light-rail implementation on labor market accessibility: a transportation equity perspective. *J. Transport Land Use* 5 (3), 28–39.
- Geurs, K.T., van Wee, B., 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. *J. Transport Geography* 12 (2), 127–140.
- Google, Inc., 2013. Transit <<http://developers.google.com/transit/>>.
- Graham, M.R., Kutzbach, M.J., McKenzie, B., 2014. Design comparison of LODES and ACS commuting data products. U.S. Census Bureau Center for Economic Studies, Working Paper CES-WP-14-38.
- Hansen, W., 1959. How accessibility shapes land use. *J. Am. Inst. Planners* 25 (2), 73–76.
- Horner, M.W., 2004. Spatial dimensions of urban commuting: a review of major issues and their implications for future geographic research. *Prof. Geogr.* 56 (2), 160–173.
- Horner, M.W., Murray, A.T., 2002. Excess commuting and the modifiable areal unit problem. *Urban Stud.* 39 (1), 131–139.
- Horner, M.W., Schleith, D., 2012. Analyzing temporal changes in land-use-transportation relationships: a LEHD-based approach. *Appl. Geogr.* 35 (1–2), 491–498.
- Ingram, D.R., 1971. The concept of accessibility: a search for an operational form. *Regional Stud.* 5 (2), 101–107.
- Kawabata, M., 2003. Job access and employment among low-skilled autoless workers in US metropolitan areas. *Environ. Plann. A* 35 (9), 1651–1668.
- Kawabata, M., 2009. Spatiotemporal dimensions of modal accessibility disparity in Boston and San Francisco. *Environ. Plann. A* 41 (1), 183.
- Kawabata, M., Shen, Q., 2007. Commuting inequality between cars and public transit: the case of the San Francisco bay area, 1990–2000. *Urban Stud.* 44 (9), 1759.
- Korkmaz, T., Kruntz, M., 2001. Multi-constrained optimal path selection. In: 20th Annual Joint Conference of the IEEE Computer and Communications Societies 2, pp. 834–843.
- Krizek, K.J., Iacono, M., El-Geneidy, A., Liao, C.-F., Johns, R., 2009. Access to destinations: application of accessibility measures for non-auto travel modes. Minnesota Department of Transportation.
- Kuipers, F., van Meigham, P., Korkmaz, T., Kruntz, M., 2002. An overview of constraint-based path selection algorithms for QoS routing. *Commun. Mag., IEEE* 40 (12), 50–55.
- Lei, T., Church, R., 2010. Mapping transit-based access: integrating GIS, routes, and schedules. *Int. J. Geogr. Inf. Sci.* 24 (2), 283–304.
- Levine, J., Gengs, J., Shen, Q., Shen, Q., 2012. Does accessibility require density or speed? A comparison of fast versus close in getting where you want to go in US metropolitan regions. *J. Am. Plann. Assoc.* 78 (2), 157–172.
- Mavoa, S., Witten, K., McCreanor, T., O'Sullivan, D., 2012. GIS based destination accessibility via public transit and walking in Auckland, New Zealand. *J. Transport Geography* 20 (1), 15–22.
- McFadden, D., 1974a. Conditional Logit Analysis of Qualitative Choice Behavior. Academic Press, pp. 105–142.
- McFadden, D., 1974b. The measurement of urban travel demand. *J. Public Econ.* 3 (4), 303–328.
- McFadden, D., Johnson, M., Reid, F., 1973. Travel Demand Forecasting Study. Institute of Urban and Regional Development, University of California.
- Menard, S., 2000. Coefficients of determination for multiple logistic regression analysis. *Am. Stat.* 54 (1), 17–24.
- Metropolitan Council, 2010. 2030 Transportation Policy Plan.
- Morris, J.M., Dumble, P., Wigan, M.R., 1979. Accessibility indicators for transport planning. *Transport. Res. Part A* 13 (2), 91–109.
- Niedzielski, M.A., Horner, M.W., Xiao, N., 2013. *Transport. Res. Part A: Policy Practice* 58, 129–143.
- Owen, A., Levinson, D., 2012. Annual accessibility measure for the Twin Cities metropolitan area. Minnesota Department of Transportation.
- Polzin, S.E., Pendyala, R.M., Navari, S., 2002. Development of time-of-day-based transit accessibility analysis tool. *Transport. Res. Rec.: J. Transport. Res. Board* 1799 (1), 35–41.

- Shen, Q., 2006. Job accessibility as an indicator of auto-oriented urban structure: a comparison of Boston and Los Angeles with Tokyo. *Environ. Plann. B: Plann. Des.* 33, 115–130.
- Spear, B.D., 2011. Improving employment data for transportation planning. American Association of State Highway and Transportation Officials.
- Taylor, B.D., Miller, D., Iseki, H., Fink, C., 2009. Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. *Transport. Res. Part A* 43 (1), 60–77.
- Tilahun, N., Levinson, D., 2011. Work and home location: possible role of social networks. *Transport. Res. Part A* 45 (4), 323–331.
- Wang, F., Xu, Y., 2011. Estimating O-D travel time matrix by Google maps API: implementation, advantages, and implications. *Ann. GIS* 17 (4), 199–209.
- Wu, B.M., Hine, J.P., 2003. A PTAL approach to measuring changes in bus service accessibility. *Transport Policy* 10 (4), 307–320.
- Zhu, S., 2010. The roads taken: theory and evidence on route choice in the wake of the I-35W Mississippi River Bridge collapse and reconstruction. PhD thesis, University of Minnesota.